



**University of  
Zurich<sup>UZH</sup>**

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2017

---

## **Molecular mechanisms of adaptation and speciation: why do we need an integrative approach?**

Byers, Kelsey J R P ; Xu, Shuqing ; Schlüter, Philipp M

**Abstract:** Understanding divergent adaptation and ecological speciation requires the synthesis of multiple approaches, including phenotypic characterization, genetics and genomics, realistic assessment of fitness and population genetic modelling. Current research in this field often approaches this problem from one of two directions: either a mechanistic approach—seeking to link phenotype, genotype and fitness, or a genomic approach—searching for signatures of divergence or selection across the genome. In most cases, these two approaches are not synthesized, and as a result, our understanding is incomplete. We argue that research in adaptation and evolutionary genetics needs to integrate these approaches for multiple reasons, including progress towards understanding the architecture and evolutionary history of adaptation and speciation loci, the ability to untangle linkage and pleiotropy, increased knowledge of mechanisms of genomic evolution and insights into parallel evolutionary events. Identifying the genetic underpinnings of adaptation and ecological speciation is not necessarily the end goal of research, but it is an integral part of understanding the evolutionary process. As a result, it is critical to utilize both genetic and genomic approaches. Challenges remain, particularly in nonmodel organisms and in our ability to synthesize results from multiple experimental systems. Nonetheless, advances in genetic and genomic techniques are increasingly available in a diverse array of systems, and the time is ripe to exploit the synthesis of these two approaches to increase our understanding of evolution.

DOI: <https://doi.org/10.1111/mec.13678>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-124154>

Journal Article

Accepted Version

Originally published at:

Byers, Kelsey J R P; Xu, Shuqing; Schlüter, Philipp M (2017). Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology*, 26(1):277-290.

DOI: <https://doi.org/10.1111/mec.13678>

Byers KJRP, Xu S, Schlüter PM (2017) Molecular mechanisms of adaptation and speciation: why do we need an integrative approach? *Molecular Ecology*, doi: 10.1111/mec.13678.

#### Molecular mechanisms of adaptation and speciation: why do we need an integrative approach?

Kelsey J.R.P. Byers<sup>1</sup>, Shuqing Xu<sup>2\*</sup>, Philipp M. Schlüter<sup>1\*</sup>

<sup>1</sup>Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, CH-8008 Zurich, Switzerland; <sup>2</sup>Max Planck Institute for Chemical Ecology, Hans-Knöll-Straße 8, D-07745, Jena, Germany.

\*To whom correspondence should be addressed: Philipp Schlüter  
<philipp.schlueter@systbot.uzh.ch>, Tel. +41 44 63 48328; or Shuqing Xu <sxu@ice.mpg.de>

**Running title:** Molecular adaptation and speciation

**Keywords:** ecological speciation, molecular adaptation, ecological genomics, integrative approach, candidate genes, genome scan

#### ABSTRACT

Understanding divergent adaptation and ecological speciation requires the synthesis of multiple approaches, including phenotypic characterization, genetics and genomics, realistic assessment of fitness, and population genetic modelling. Current research in this field often approaches this problem from one of two directions – either a mechanistic approach, seeking to link phenotype, genotype, and fitness – or a genomic approach, searching for signatures of divergence or selection across the genome. In most cases these two approaches are not synthesized, and as a result our understanding is incomplete. We argue that research in adaptation and evolutionary genetics needs

23 to integrate these approaches for multiple reasons, including progress towards understanding the  
24 architecture and evolutionary history of adaptation and speciation loci, the ability to untangle linkage  
25 and pleiotropy, increased knowledge of mechanisms of genomic evolution, and insights into parallel  
26 evolutionary events. Identifying the genetic underpinnings of adaptation and ecological speciation is  
27 not necessarily the end goal of research, but it is an integral part of understanding the evolutionary  
28 process. As a result, it is critical to utilize both genetic and genomic approaches. Challenges remain,  
29 particularly in non-model organisms and in our ability to synthesize results from multiple  
30 experimental systems. Nonetheless, advances in genetic and genomic techniques are increasingly  
31 available in a diverse array of systems, and the time is ripe to exploit the synthesis of these two  
32 approaches to increase our understanding of evolution.

## INTRODUCTION

The concept of ecological speciation intimately links divergent selection and adaptation to the origin of different species. Divergent adaptation and ecological speciation (hereafter A&S) are the consequences of divergent selection imposed by the local environment, population demography, and genetic architecture of traits that affect organismal fitness (Schluter 2001; Nosil 2012). Although much progress has been made in each of these areas, particularly due to recent sequencing advances, fewer attempts have been made to integrate these areas in one study system. This can be seen from the often divided A&S research programmes: a) study of the molecular mechanisms of A&S, which focuses on identifying genetic and biochemical mechanisms underlying adaptation; and b) ecological speciation/adaptation genomics, which focuses on how natural selection drives divergence of specific genomic loci. While both molecular adaptation and speciation genomics approaches have provided valuable insight into the genetics underlying A&S, these two approaches are rarely combined: molecular adaptation studies of targeted loci are rarely expanded to a genome-wide and population scale, and genome scans for signatures of selection and divergence are rarely followed up by in-depth molecular work. Below, we argue that these two research programmes have their own strengths and limitations, but are complementary to each other in studying A&S. Therefore, we propose that integration of these two research areas is needed for a more complete and unbiased understanding of the processes of A&S. In addition, we identify the key challenges and opportunities that will arise for integrating approaches to the study of molecular and genomic A&S (Figure 1).

## MOLECULAR ADAPTATION AND SPECIATION GENOMICS

The modern synthesis, in uniting genetics and evolution, suggested that evolutionary processes could have a known genetic basis and provided an impetus to identify the molecular basis of A&S (Kutschera & Niklas 2004). Today, the molecular study of adaptation ranges more broadly,

including diverse molecular mechanisms of evolution at multiple levels. We consider this molecular study to include not just genes, but also *e.g.* transcriptional regulation of gene networks, biochemical pathways, the role of small RNAs, and mechanisms by which environmental input is processed. Importantly, mechanistic study does not stop once “a gene” has been found: we would like to know the effects, origins, and interactions of different alleles and mutations (Barrett & Hoekstra 2011). In seeking to understand molecular adaptation, we strive to understand the interaction of genetics and evolution in both directions: first, what is the genetic basis underlying an organism’s adaptation to its environment, and second, how do ecological and evolutionary forces act upon the organism’s genetic information to result in adaptation?

Our approach to understanding molecular adaptation often consists of the following steps: a) identifying adaptive traits which increase organismic fitness in various environments; b) identifying the genetic basis of these adaptive traits, typically using genetic mapping or a candidate gene approach; and c) measuring the fitness effects of variation in the identified candidate genes, alleles, or mutations. This forward genetics approach, linking genotype-phenotype-fitness, represents the gold standard for studying molecular mechanisms of A&S, and tremendous progress is being made in applying this approach in both model and non-model systems (Barrett & Hoekstra 2011; Watt 2013).

Despite clear advantages of this approach, there are several limitations in each individual step and their integration. First, identifying truly adaptive traits remains challenging. As human beings, we are biased towards studying traits accessible to human senses (*e.g.* flower or coat colour), while neglecting those that are invisible or inaccessible to us, but may be more appealing to selective agents (*e.g.* UV patterning in *Petunia* flowers, Sheehan *et al.* 2016, or surface hydrocarbons on orchid flowers, Xu *et al.* 2012b, both attractive to insect pollinators). Additionally, natural selection on traits is often a dynamic process (Siepielski *et al.* 2009), thus selection coefficient estimation requires multi-year measurements and may not represent the true selective regime that shaped the evolution of

the traits in the past. The effects of a trait on fitness may vary at different life history stages and can be difficult to measure in many systems (*e.g.* survivorship in long-lived organisms) (Figure 1: Challenge 1). Second, identifying the genetic architecture of adaptive traits using a mapping-based approach can be biased. The effect of genes on phenotypic traits can be highly dependent on genetic, genomic, and environmental background (Beavis 1998; Otto & Jones 2000) (Figure 1: Challenge 2), and thus the identified candidate genes associated with an adaptive trait might be dependent on the choice of mapping population and environment. Mapping with artificial populations has the advantage of tractability, but in the forward genetics approach, mapping techniques are often applied outside of the natural environmental context, leading to potential bias in applying mapping results to ecological functions. Additionally, these experiments are rarely followed up for evidence of selection on identified candidates. Gene functions and selection within the present environment and time may not reflect *historical* selection, gene function, and previous fluctuating genomic and environmental context, and this context may be impossible to determine (Figure 1: Challenge 1). As a result, it may be more useful to study incipient speciation events where lineage separation is not yet complete (Via 2009; Stankowski & Streisfeld 2015). In many cases, determining when adaptation occurred is itself a challenging task (but see Danley & Kocher 2001). Third, quantitative measurement of the fitness effects of variation in candidate genes is both challenging and time consuming. In the case of speciation, understanding the direct effects of candidate genes on gene flow is important (Hendry & Taylor 2004; Morjan & Rieseberg 2004), but this can be difficult to quantify (Figure 1: Challenge 3). It is also possible that selection may act on genes that have small phenotypic but large fitness effects, rather than more easily detectable genes with large phenotypic but small fitness effects. Finally, when integrating all information in one system, the same natural populations or those with a similar population makeup should be used for all three steps, since the fitness effects of traits and genes are highly dependent on their genetic and environmental background (Barrett & Hoekstra 2011); this can be especially challenging for biologically interesting non-model organisms.

In contrast to the forward genetics approach, the genomic adaptation approach (“population genomics” *sensu* Stinchcombe & Hoekstra 2008) aims to directly identify putative adaptive loci in the genome by searching for molecular signatures of selection between populations or closely related species (akin to a reverse genetics approach). Rather than considering the molecular basis of adaptation in just the context of one or a few genes or pathways, an ecological speciation/adaptation genomics approach allows us to query the genomic landscape of these processes; this process is aided by recent advances in “omics” approaches, including the relative ease of sequencing and assembling genomes or transcriptomes of different populations and the ability to scan the diverse array of proteins and metabolites produced in specific conditions. The advantage of this approach is the ability to directly search for end products (or footprints) resulting from the A&S process, and it also provides an unbiased way of detecting loci that are under selection. For example, many genomic loci might be under selection due to specific environmental conditions or agents of selection that cannot be measured or replicated in current experimental setups (MacColl 2011), thus will likely not be identified through the forward genetics phenotype-genotype-fitness approach. Insights into historical population phenomena such as admixture and assortative mating are also possible with a genome-wide approach (Lexer *et al.* 2010). Furthermore, this approach also allows us to detect change above the gene level – *e.g.* recombination or inversion – which may be the main genetic basis of evolution in a variety of systems, as some authors have suggested (Coyne & Orr 2004; Fishman *et al.* 2013).

Currently, searching for loci under selection is often carried out by scanning the genome for outliers using Wright’s fixation index ( $F_{ST}$ , Wright 1951; Beaumont 2005), but other metrics exist. These include Tajima’s D (Tajima 1989), the Hudson-Kreitman-Aguadé (HKA) test (Hudson *et al.* 1987), or  $d_N/d_S$  ( $\omega$ ) ratios (*e.g.* Torrents *et al.* 2003), each of which comes with certain assumptions and potential pitfalls (Charlesworth 1998; Hughes 2007; Nei *et al.* 2010; Cruickshank & Hahn 2014). To address some of these issues, other indices have recently been developed, often based on  $F_{ST}$

(Meirmans & Hedrick 2011; Pannell & Fields 2014). These analyses often identify putative “islands” or “continents” of speciation, the size and number of which can be informative in understanding the process of speciation in a given system (Feder *et al.* 2012).

However, the advantage of directly searching for end products of A&S at the same time also leads to several limitations. In addition to selection, several other factors can also result in similar statistical patterns at genomic loci, such as population demographic changes, selective sweeps, and divergence hitchhiking (Gompert *et al.* 2012; Via 2012) (Figure 1: Challenge 4). Although optimized experimental design and different advanced population genomic modelling approaches have been developed to disentangle these different confounding effects (Cutter & Payseur 2013), it remains extremely challenging to identify genomic loci that contributed to local adaptation based only on population genomic studies (Flaxman *et al.* 2013), especially among genes located in a recent selective sweep region. Furthermore, many genomic loci that show signatures of selection may not have any known or substantiated functional annotations, especially in non-model organisms (Figure 1: Challenge 5). This prevents further study of why and how those loci contribute to the A&S process, which is often the primary goal of evolutionary studies.

## **BENEFITS OF AN INTEGRATIVE APPROACH**

These molecular and genomic approaches are complementary to each other, and integrating them in one system will provide a more complete understanding of A&S processes as well as offering context to these approaches. The idea of combining molecular and genomic approaches has been proposed previously (Stinchcombe & Hoekstra 2008), but research doing so has so far been limited to only a few systems (see examples in Box 1C). For much of the research on the molecular basis of adaptation, which is often limited to a single-gene context, adding an adaptation genomic view will provide greater mechanistic understanding of how and why a given gene was (or was not) under selection (Box 1A). For example, a gene identified via linkage or association mapping may interact in a



155 regulatory or biosynthetic network with other genes in the genome, or may share common  
156 regulatory elements with linked loci – in other words, the genomic context may be critical to  
157 understanding the gene's function (Dewey 2011). Such linked genes may contribute to the process of  
158 A&S, and will be missed with a focused molecular approach, whereas a selective sweep may be  
159 picked up with a genome-wide scan. Furthermore, combining molecular adaptation and genomics  
160 may facilitate rapid identification of causal mutations that contribute to adaptation. Traditional  
161 mapping and more recent genome-wide approaches to the identification of genes underlying  
162 adaptive traits both often result in a large number of candidate genes. Functionally testing each  
163 individual gene is costly and time consuming, yet it is only by assessing the function of the most  
164 promising individual genes that we can make firm conclusions about their contribution to adaptation,  
165 and the physical properties that favoured specific variants (Harms & Thornton 2013). As a result, we  
166 encourage researchers and funding agencies to allocate resources to functional testing of highly  
167 significant candidate genes. In an ideal world, this would consist of functional testing within the  
168 system in question – *e.g.* via transgenesis or targeted mutation, or less precisely via high-resolution  
169 near-isogenic lines, targeting either coding or regulatory candidate sequences. Ideally this would be  
170 done in both directions to verify that candidate alleles are both necessary and sufficient to produce  
171 the phenotypic and selective effects. Currently, some questions are amenable to functional testing  
172 via techniques such as heterologous expression, although this carries some risks in interpretation  
173 (Kramer 2015). Other, detailed, molecular techniques can provide supporting evidence, but by  
174 themselves are inadequate to verify a candidate gene's function *in vivo*. With any functional test,  
175 measuring the effects on a variety of phenotypes – not just the trait in question – will be critical to  
176 assessing the full effect of a specific allele on the organism. Integrating an adaptation genomic  
177 approach may help to rapidly narrow down the focus to a small set of key genes that likely were  
178 targets of selection. For example, Smadja *et al.* 2012 performed  $F_{ST}$  and  $\pi_a/\pi_s$  tests on a set of 153  
179 candidate genes potentially underlying pea aphid speciation, and found 19 genes showing increased

divergence across three populations; a larger set of genes and aphid populations has confirmed and expanded on this work (Eyres *et al.* 2017, this issue). Moreover, A&S genomics can also provide further insights on population demographic change, which is an important factor that influences the process of A&S (Städler *et al.* 2005).

Similarly, ecological speciation genomics benefits from incorporating a more focused molecular view. Many current analyses of speciation genomics stop at the genome scan or annotation level, which may result in markers potentially linked to A&S, but not the actual genetic underpinnings of these processes or the identification of specific candidate speciation genes (if they exist) (Box 1B). In our opinion, there are a number of new insights that can be gained from identifying the molecular functions of candidate genes associated with A&S processes. a) it can help to rule out false positives. Genome-wide scans often result in a large number of false positives which prevent direct conclusions from being drawn (Lotterhos & Whitlock 2014). In particular, tightly linked genes may have similar signatures of differentiation, even if only one was the actual target of selection. Identifying the molecular function of these tightly linked genes and testing their fitness effects will help to identify the true targets of selection. By narrowing our focus down from the genomic region to individual genes and regulatory elements, it becomes easier to see whether the region likely contains legitimate targets of selection, rather than trusting the result of genome-wide scans alone or choosing a “likely-looking” candidate gene and failing to verify its function. b) it will allow us to understand whether complex evolutionary scenarios are due to the linkage of several necessary functions (*e.g.* via linkage disequilibrium) or due to pleiotropy of a single gene (Wagner & Zhang 2011; Paaby & Rockman 2013). c) it can help us to understand the most significant driving forces that have shaped recent A&S processes. Loci showing signatures of selection often either have unknown functions or are only functionally characterized in model organisms. Identifying the exact molecular functions and the traits that the target gene is associated with can provide insight on which traits were targets of selection and what the key driving forces were. d) it will allow us to identify and

analyse parallel evolution events (Colosimo *et al.* 2005; Rausher 2008; Manceau *et al.* 2010; Streisfeld & Rausher 2011; Conte *et al.* 2012). e) it will provide new insights into the trajectory of A&S, which may often be associated with the likelihood and direction of trait transitions. This can be seen from the fact that trait transitions that involve gene or pathway degeneration caused by accumulation of loss-of-function mutations are likely to exist in a one-way direction only (*e.g.* Rausher 2008; Streisfeld & Rausher 2011). Therefore, identifying causal genes and mutations will allow modelling of the direction of A&S events (*e.g.* Weinreich *et al.* 2006; Franke *et al.* 2011; Xu & Schlüter 2015). f) it will allow us to identify other factors, such as genome evolution events (*e.g.* gene conversion, whole genome duplication, and tandem duplications) that may influence the process of A&S (Volff 2005). g) it will allow us to more broadly understand the molecular nature of speciation loci, *e.g.* whether duplication allows evolutionary innovations, what effect sizes are necessary to promote the speciation process, or whether certain gene classes are over- or underrepresented in the speciation process. Until we understand the molecular genetics behind A&S, it will be difficult to draw conclusions in these areas. Although the identification of the specific molecular basis of A&S is not necessarily the end goal in any given system (Rockman 2012), a complete understanding of the process of A&S requires that we understand its genetic underpinnings, and we argue that this requires an integrative approach.

## CHALLENGES OF INTEGRATION

Combining genome-wide approaches with focused molecular adaptation approaches remains a challenge (Figure 1: Challenge 7), particularly in systems with limited genetic resources and those difficult to rear in controlled conditions. The current dearth of examples and study systems combining these approaches is certainly not due to a lack of interest; rather, in many systems we are only now beginning to have the resources necessary to do so. Barrett and Hoekstra (2011) suggested

an approach for designating an allele as adaptive, with three contributing lines of evidence: phenotype, genotype, and fitness. In their Figure 1, “Genome-wide scans for signatures of selection” (fitness and genotype) does not overlap with “Genetic mapping of phenotypic traits” (genotype and phenotype) – the separation of molecular adaptation and genome-wide approaches is evident. On the one hand, focused molecular adaptation approaches yield strong data on the genetic underpinnings of phenotypes of interest, but may reflect a bias for the most tractable or “interesting” phenotypes (Rockman 2012; but see Lee *et al.* 2014). Unbiased genotype-phenotype approaches such as quantitative trait locus (QTL) mapping exist, but establishing the necessary mapping populations is difficult to impossible in many non-model systems. On the other hand, genome-wide association studies (GWAS; Visscher *et al.* 2012) for phenotype-genome association are beginning to be used in some non-model systems (Hecht *et al.* 2013). In particular, techniques developed for working with human genetics may benefit non-model systems with similar constraints (e.g. the inability to establish mapping or common garden populations and difficulties in functional verification); by applying these techniques to natural populations, we can overcome some of the environmental and genomic biases inherent in some traditional mapping approaches. With any method, caution should be taken when choosing loci to focus on, as this can suffer from bias (Wayne & McIntyre 2002).

Understanding the molecular functions of candidate genes from either the molecular adaptation or the ecological genomics approach remains a serious problem, particularly as many non-model systems lack the extensive genetic and laboratory resources necessary to tackle these issues; however, as with the development of genetic resources, techniques for studying candidate gene functions are continuously improving. Choice of study system plays a large role here, while at the same time limiting research to tractable systems may not reflect the majority of biological reality. A particular issue is that it is unclear to what extent molecular mechanisms are limited to the populations or species under study (Gabaldón & Koonin 2013; Kramer 2015). Fortunately, with recent

advances in sequencing techniques, as well as analysis tools developed for work with model system data, the number of amenable systems is increasing and choice can increasingly be made on biological interest rather than tractability (Ekblom & Galindo 2011); in this way, many systems with established natural history data have become “emerging model systems” (Abzhanov *et al.* 2008). The availability of a larger number of emerging model systems also facilitates the genetic study of repeated evolutionary events, thus allowing discussion of the “repeatability of evolution” (*e.g.* Colosimo *et al.* 2005; Rausher 2008; Manceau *et al.* 2010; Streisfeld & Rausher 2011; Conte *et al.* 2012). In practice, integrating ecological genomics into molecular adaptation in non-model systems will require familiarity with next-generation sequencing and analysis techniques as well as funding for these initiatives, but the ongoing development of accessible genomic analysis tools and sequencing centres has greatly improved the prospects for this integration.

At the same time, it can be difficult to combine experimental results from disparate systems into a conceptual whole (Figure 1: Challenge 6): differing ecological and evolutionary contexts, genetic and developmental processes, and ecosystem function hinder our ability to compare results in this case. As a particular example of this problem, utilizing tools such as homology-based gene identification and gene ontology categorization can be risky in non-model systems (van den Berg *et al.* 2009; Tagu *et al.* 2014; Kramer 2015), and experimental assessment of function is crucial to fully understand the role of genotypes in adaptation in a given system. A critical part of functional validation is the documentation and reporting of situations in which the results demonstrate divergent functionalities (including loss of function) in non-model systems from those predicted using prior data; at present, this reporting likely suffers from a bias against publishing seemingly negative results. Recent reviews have begun to assemble conclusions from a variety of systems (*e.g.* Hoekstra 2006; Lowry *et al.* 2008; Martin & Orgogozo 2013; Groot *et al.* 2016), showing the promise of integrating results across non-model systems.

In the emphasis on understanding the genetic and genomic basis of phenotypes, we often fail to integrate ecology and natural history, limiting our conclusions to solely lab-based *e[x] natura* experiments and thus missing their ecological and evolutionary significance (Shimizu *et al.* 2011). Assessing the fitness or isolation effects of candidate genes in an ecological context in the field (*i.e. in natura*) also remains a serious challenge, particularly with genetically modified organisms; in some cases, understanding the genetic basis of a phenotype may be combined with manipulative experiments to test selection on that specific phenotype (*e.g.* Xu *et al.* 2012a). Different drivers of divergent adaptation (*e.g.* abiotic and biotic factors) and different speciation contexts (*e.g.* allopatry versus sympatry) also will change the significance of a given phenotype and genetic/genomic combination. During early phases of divergence, when selection possibly directly acts on few genes (Feder *et al.* 2012), the relative importance of these genes for species divergence is increased. Therefore, biochemical and molecular constraints to the evolution at these focal loci may become directly important for the process of speciation by facilitating or constraining divergence. In phenotypes that appear to be under simpler genetic control, integrating ecological genomics may be less critical, although genomic techniques (*e.g.* genome-wide scans) can help to test whether the phenotype is indeed as genetically simple as hypothesized, and population genomics can inform us as to the complexity of the genetic basis of adaptation.

## OUTLOOK

Fortunately, combining ecological speciation genomics with molecular adaptation in a rigorous fashion (Figure 1: Challenge 7) has become more feasible with the decrease in sequencing costs over time (Ekblom & Galindo 2011). A bidirectional approach may be particularly powerful: using genome-wide scans to identify loci under selection in natural populations and combining this knowledge with more focused molecular knowledge from linkage and association mapping (Stinchcombe & Hoekstra 2008; Via *et al.* 2012; Li *et al.* 2013). In systems not amenable to traditional

mapping techniques, there may be the option to harness natural variation using GWAS and similar approaches originally developed for use in human genetics. Ideally, quantitative trait measurement and assessment of fitness should happen in a realistic field environment using translocation experiments where feasible (*e.g.* Angert & Schemske 2005; Knight *et al.* 2006; Leinonen *et al.* 2009). Phylogenetic ancestral reconstruction and model-based approaches may provide some insight into the historic genetic and environmental context of A&S (*e.g.* Gaucher *et al.* 2008). The most powerful research in A&S genetics and genomics will combine all relevant factors – phenotype, genotype, fitness consequences, population genomics, and evolutionary history – in a single system (Box 1C).

One potential approach might be as follows: first, for the system in question, identification of what factors to focus on is critical. Although the vast diversity of phenotypic traits is a fundamentally attractive part of biology, understanding which traits underlie adaptation is crucial to understanding how this diversity arose. For example, what are the most significant stages of reproductive isolation, or what environmental factors are driving adaptation (Ramsey *et al.* 2003; Widmer *et al.* 2009)? Second, identify which regions of the genome are under divergent selection; this and subsequent mapping steps can use the same set of markers to save costs. It may also be helpful to identify the frequency and pattern of genome-wide single nucleotide polymorphisms (SNPs), recombination, gene conversion, and chromosomal re-arrangements for the studied system. Third, consider traditional linkage mapping or GWAS approaches to narrow down the genetic basis of phenotypic traits that seem to be under selection or are involved in reproductive isolation. Fourth, combine the genomic information from the previous two steps to look for overlaps between candidate loci for a given trait and genomic regions under selection (*e.g.* Via *et al.* 2012; Li *et al.* 2013). Finally, where reasonable evidence exists that a specific genomic region is under selection and involved in the phenotype of interest, consider thorough genetic and biochemical characterization of the locus in order to understand the molecular mechanisms of genes that are under divergent selection. Where possible, one should also measure the quantitative fitness consequences of different alleles, or

ideally, specific mutations (Barrett & Hoekstra 2011). Other approaches in existing studies include investigating signatures of selection among populations for existing candidate genes (Box 1). If the system allows, using CRISPR (Cong *et al.* 2013) or other transgenic techniques to identify the putative function and possible fitness effects of these genes would be ideal, but for many systems this may not be possible. In this case, other options include studying the function of candidate genes in closely related model species or characterizing gene co-expression networks to identify the major functions of genes that are co-expressed with the candidate genes (Higashi & Saito 2013). These approaches will yield an in-depth understanding of ecological A&S genetics.

## **CONCLUSION**

Adaptation and ecological speciation are dynamic processes that are shaped by ecological conditions as well as the genetic makeup of the organism and its evolutionary history (Nosil & Feder 2012). Understanding the mechanisms underlying these complex processes is challenging and requires an integrative approach. Molecular adaptation studies usually provide an episodic, mechanistic view of how A&S may have occurred, and A&S genomics allows us a view of the evolutionary end point. Combining these two complementary approaches will not only provide a more complete and unbiased understanding of the entire process, but may allow us to compare the mechanisms of A&S among systems and to draw conclusions about the molecular nature of A&S loci and their evolution (Figure 1: Challenge 7). We believe that research programmes that integrate tools from field ecology, molecular biology, population genomics, and individual-based experimentally informed modelling are needed to study A&S (see Lazebnik 2002 for an inspiring read). Systems that have experienced recent speciation and adaptation events – particularly those where adaptation and speciation are presently ongoing – are extremely valuable for understanding how and why A&S occurred, since underlying environmental factors are better known and we may be better able to elucidate the initial processes shaping A&S (Via 2009). For example, organisms that have adapted to



recent anthropogenic changes (*e.g.* agriculture and mining), or speciation events triggered by recent climatic changes such as glacial recession, can serve as extremely valuable systems to systematically investigate A&S. Furthermore, studies of the process of adaption using both forward genetics and genome-wide approaches in experimentally evolved systems will provide a more complete conceptual understanding of adaptation and ecological speciation. Full knowledge of the genetic substrate upon which selection by the environment ultimately works is crucial if we want to gain a complete understanding of evolutionary change.

## ACKNOWLEDGEMENTS

The authors would like to thank the organizers and participants of the European Society for Evolutionary Biology's 2015 meeting in Lausanne, Switzerland. In particular, we thank the participants of the ESEB symposium "The molecular basis of adaptation and ecological speciation" for stimulating discussion and the Swiss National Science Foundation (SNF) for financial support for this symposium under grant 31CO30\_160611. KJRPB is funded by a PLANT FELLOWS Postdoctoral Fellowship (European Union: FP7-PEOPLE-2010-COFUND Proposal No. 267243) and the University of Zurich. SX acknowledges funding by the SNF, grant PEBZP3-142886, a Marie Curie Intra-European Fellowship (IEF), grant 328935, and the Max Planck Society. PMS is funded by the SNF, grant 31003A\_155943, and the University of Zurich. The authors also thank Sean Rogers, two anonymous reviewers, and Roman Kellenberger for comments on the manuscript, as well as Aaron Comeault and Moritz Muschick for kindly providing images of *Timema*.

## AUTHOR CONTRIBUTIONS

KJRPB, SX and PMS contributed ideas; KJRPB and SX wrote the manuscript; KJRPB, SX and PMS revised the manuscript.

## DATA ACCESSIBILITY

Not applicable.

## REFERENCES

- Abzhanov A, Extavour CG, Groover A *et al.* (2008) Are we there yet? Tracking the development of new model systems. *Trends in Genetics*, **24**, 353-360.
- Angert AL, Schemske DW (2005) The evolution of species' distributions: reciprocal transplants across the elevational ranges of *Mimulus cardinalis* and *M. lewisii*. *Evolution*, **59**, 1671-1684.
- Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics*, **12**, 767-780.
- Beaumont MA (2005) Adaptation and speciation: what can  $F_{ST}$  tell us. *Trends in Ecology and Evolution*, **20**, 435-440.
- Beavis WD (1998) in Molecular dissection of complex traits (ed. Paterson AH), 145-162. CRC Press, New York City, New York.
- Bélanger-Deschênes S, Couture P, Campbell PGC *et al.* (2013) Evolutionary change driven by metal exposure as revealed by coding SNP genome scan in wild yellow perch (*Perca flavescens*). *Ecotoxicology*, **22**, 938-957.
- Bradshaw HD Jr., Otto KG, Frewen BE *et al.* (1998) Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). *Genetics*, **149**, 367-382.
- Bradshaw HD Jr., Schemske DW (2003) Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature*, **426**, 176-178.
- Bradshaw HD Jr., Wilbert SM, Otto KG *et al.* (1995) Genetic mapping of floral traits associated with reproductive isolation in monkeyflowers (*Mimulus*). *Nature*, **376**, 762-765.
- Byers KJRP, Vela JP, Peng F *et al.* (2014) Floral volatile alleles can contribute to pollinator-mediated reproductive isolation in monkeyflowers (*Mimulus*). *The Plant Journal*, **80**, 1031-1042.
- Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538-543.
- Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*, **307**, 1928-1933.

405 Comeault AA, Flaxman SM, Riesch R *et al.* (2015a) Selection on a genetic polymorphism counteracts  
 406 ecological speciation in a stick insect. *Current Biology*, **25**, 1975-1981.

407 Comeault AA, Ferreira C, Dennis S *et al.* (2015b) Evolution of color phenotypes in two distantly  
 408 related species of stick insect: different ecological regimes acting on similar genetic architectures.  
 409 *bioRxiv*, 023481.

410 Cong L, Ran FA, Cox D *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*,  
 411 **339**, 819-823.

412 Conte GL, Arnegard ME, Peichel CL *et al.* (2012) The probability of genetic parallelism and  
 413 convergence in natural populations. *Proceedings of the Royal Society of London B: Biological Sciences*,  
 414 **279**, 5039-5047.

415 Couture P, Pyle G (2008) Live fast and die young: metal effects on condition and physiology of wild  
 416 yellow perch from along two metal contamination gradients. *Human and Ecological Risk Assessment: An International Journal*, **14**, 73-96.

418 Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Sunderland, Massachusetts.

419 Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to  
 420 reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133-3157.

421 Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity  
 422 among species. *Nature Reviews Genetics*, **14**, 262-274.

423 Danley PD, Kocher TD (2001) Speciation in rapidly diverging systems: lessons from Lake Malawi.  
 424 *Molecular Ecology*, **10**, 1075-1086.

425 Dewey CN (2011) Positional orthology: putting genomic evolutionary relationships into context.  
 426 *Briefings in Bioinformatics*, **12**, 401-412.

427 Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-  
 428 model organisms, *Heredity*, **107**, 1-15.

429 Eyres I, Duvaux L, Gharbi K *et al.* (2017) Targeted re-sequencing confirms the importance of  
 430 chemosensory genes in aphid host race differentiation. *Molecular Ecology*, **THIS ISSUE**.

431 Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**,  
 432 342-350.

433 Fishman L, Stathos A, Beardsley PM *et al.* (2013) Chromosomal rearrangements and the genetics of  
 434 reproductive barriers in *Mimulus* (monkeyflowers). *Evolution*, **67**, 2547-2560.

435 Flaxman SM, Feder JL, Nosil P (2013) Genetic hitchhiking and the dynamic buildup of genomic  
 436 divergence during speciation with gene flow. *Evolution*, **67**, 2577-2591.

437 Franke J, Klözer A, De Visser JAGM *et al.* (2011) Evolutionary accessibility of mutational pathways.  
 438 *PLoS Computational Biology*, **7**, e1002134.

439 Gabaldón T, Koonin EV (2013) Functional and evolutionary implications of gene orthology. *Nature Reviews Genetics*, **14**, 360-366.

441 Gaucher EA, Govindarajan S, Ganesh OK (2008) Palaeotemperature trend for Precambrian life  
 442 inferred from resurrected proteins. *Nature*, **451**, 704-707.

443 Gompert Z, Parchman TL, Buerkle CA (2012) Genomics of isolation in hybrids. *Philosophical  
 444 Transactions of the Royal Society of London B: Biological Sciences*, **367**, 439-450.

445 Groot AT, Dekker T, Heckel DG (2016) The genetic basis of pheromone evolution in moths. *Annual  
 446 Review of Entomology*, **61**, 99-117.

447 Harms MJ, Thornton JW (2013) Evolutionary biochemistry: revealing the historical and physical  
 448 causes of protein properties. *Nature Reviews Genetics*, **14**, 559-571.

449 Hecht BC, Campbell NR, Holecek DE *et al.* (2013) Genome-wide association reveals genetic basis for  
 450 the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**,  
 451 3061-3076.

452 Hendry AP, Taylor EB (2004) How much of the variation in adaptive divergence can be explained by  
 453 gene flow? An evaluation using lake-stream stickleback pairs. *Evolution*, **58**, 2319-2331.

454 Hermann K, Klahre U, Moser M *et al.* (2013) Tight genetic linkage of prezygotic barrier loci creates a  
 455 multifunctional speciation island in *Petunia*. *Current Biology*, **23**, 873-877.

456 Higashi Y, Saito K (2013) Network analysis for gene discovery in plant-specialized metabolism. *Plant,  
 457 Cell & Environment*, **36**, 1597-1606.

458 Hoballah ME, Gübitz T, Stuurman J *et al.* (2007) Single gene-mediated shift in pollinator attraction in  
 459 *Petunia*. *Plant Cell*, **19**, 779-790.

460 Hoekstra HE (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates.  
 461 *Heredity*, **97**, 222-234.

462 Hoekstra HE, Hirschmann RJ, Bunday RA *et al.* (2006) A single amino acid mutation contributes to  
 463 adaptive beach mouse color pattern. *Science*, **313**, 101-104.

464 Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide  
 465 data. *Genetics*, **116**, 153-159.

466 Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive  
 467 selection at the nucleotide sequence level. *Heredity*, **99**, 364-373.

468 Klahre U, Gurba A, Hermann K *et al.* (2011) Pollinator choice in *Petunia* depends on two major  
 469 genetic loci for floral scent production. *Current Biology*, **21**, 730-739.

470 Knight CA, Vogel H, Kroymann J *et al.* (2006) Expression profiling and local adaptation of *Boechera  
 471 holboellii* populations for water use efficiency across a naturally occurring water stress gradient.  
 472 *Molecular Ecology*, **15**, 1229-1237.

473 Kramer EM (2015) A stranger in a strange land: the utility and interpretation of heterologous  
 474 expression. *Frontiers in Plant Science*, **6**, 734.

475 Kronforst MR, Papa R (2015) The functional basis of wing patterning in *Heliconius* butterflies: the  
 476 molecules behind mimicry. *Genetics*, **200**, 1-19.

477 Kutschera U, Niklas KJ (2004) The modern theory of biological evolution: an expanded synthesis.  
 478 *Naturwissenschaften*, **91**, 255-276.

479 Lazebnik Y (2002) Can a biologist fix a radio? - Or, what I learned while studying apoptosis. *Cancer*  
 480 *Cell*, **2**, 179-182.

481 Lee YW, Gould BA, Stinchcombe JR (2014) Identifying the genes underlying quantitative traits: a  
 482 rationale for the QTN programme. *Annals of Botany Plants*, **6**, plu004.

483 Leinonen PH, Sandring S, Quilot B *et al.* (2009) Local adaptation in European populations of  
 484 *Arabidopsis lyrata* (Brassicaceae). *American Journal of Botany*, **96**, 1129-1137.

485 Lexer C, Joseph JA, van Loo M *et al.* (2010) Genomic admixture analysis in European *Populus* spp.  
 486 reveals unexpected patterns of reproductive isolation and mating. *Genetics*, **186**, 699-712.

487 Li H, Peng Z, Yang X *et al.* (2013) Genome-wide association study dissects the genetic architecture of  
 488 oil biosynthesis in maize kernels. *Nature Genetics*, **45**, 43-50.

489 Linnen CR, Kingsley EP, Jensen JD *et al.* (2009) On the origin and spread of an adaptive allele in deer  
 490 mice. *Science*, **325**, 1095-1098.

491 Linnen CR, Poh Y-P, Peterson BK *et al.* (2013) Adaptive evolution of multiple traits through multiple  
 492 mutations at a single gene. *Science*, **339**, 1312-1316.

493 Lorenz-Lemke AP, Mäder G, Muschner VC *et al.* (2006) Diversity and natural hybridization in a highly  
 494 endemic species of *Petunia* (Solanaceae): a molecular and ecological analysis. *Molecular Ecology*, **15**,  
 495 4487-4497.

496 Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization  
 497 on the performance of  $F_{ST}$  outlier tests. *Molecular Ecology*, **23**, 2178-2192.

498 Lowry DB, Modliszewski JL, Wright KM *et al.* (2008) The strength and genetic basis of reproductive  
 499 isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society of London B:*  
 500 *Biological Sciences*, **363**, 3009-3021.

501 MacColl ADC (2011) The ecological causes of evolution. *Trends in Ecology & Evolution*, **26**, 514-522.

502 Mallarino R, Linden T, Linnen C *et al.* (2017) The role of isoforms in the evolution of cryptic coloration  
 503 in *Peromyscus* mice. *Molecular Ecology*, **THIS ISSUE**.

504 Manceau M, Domingues VS, Linnen CR *et al.* (2010) Convergence in pigmentation at multiple levels:  
 505 mutations, genes and function. *Philosophical Transactions of the Royal Society of London B: Biological*  
 506 *Sciences*, **365**, 2439-2450.

507 Martin A, McCulloch KJ, Patel NH *et al.* (2014) Multiple recent co-options of Optix associated with  
 508 novel traits in adaptive butterfly wing radiations. *EvoDevo*, **5**, 7.

509 Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalog of genetic hotspots of  
 510 phenotypic variation. *Evolution*, **67**, 1235-1250.

511 Meirmans PG, Hedrick PW (2011) Assessing population structure:  $F_{ST}$  and related measures.  
 512 *Molecular Ecology Resources*, **11**, 5-18.

513 Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and  
514 selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341-1356.

515 Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius*  
516 butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal*  
517 *Society of London B: Biological Sciences*, **367**, 343-353.

518 Nei M, Suzuki Y, Nozawa M (2010) The neutral theory of molecular evolution in the genomic era.  
519 *Annual Review of Genomics and Human Genetics*, **11**, 265-289.

520 Nosil P (2007) Divergent host plant adaptation and reproductive isolation between ecotypes of  
521 *Timema cristinae* walking sticks. *The American Naturalist*, **169**, 151-162.

522 Nosil P (2012) Ecological speciation. Oxford University Press, New York City, New York.

523 Nosil P, Crespi BJ, Sandoval CP (2002) Host-plant adaptation drives the parallel evolution of  
524 reproductive isolation. *Nature*, **417**, 440-443.

525 Nosil P, Feder JL (2012) Genomic divergence during speciation: causes and consequences.  
526 *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**, 332-342.

527 Otto SP, Jones CD (2000) Detecting the undetected: estimating the total number of loci underlying a  
528 quantitative trait. *Genetics*, **156**, 2093-2107.

529 Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends in Genetics*, **29**, 66-73.

530 Pannell JR, Fields PD (2014) Evolution in subdivided plant populations: concepts, recent advances and  
531 future directions. *New Phytologist*, **201**, 417-432.

532 Ramsey J, Bradshaw HD Jr., Schemske DW (2003) Components of reproductive isolation between the  
533 monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*, **57**, 1520-1534.

534 Rausher MD (2008) Evolutionary transitions in floral color. *International Journal of Plant Sciences*,  
535 **169**, 7-21.

536 Reed RD, Papa R, Martin A *et al.* (2011) *optix* drives the repeated convergent evolution of butterfly  
537 wing pattern mimicry. *Science*, **333**, 1137-1141.

538 Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does  
539 not glitter. *Evolution*, **66**, 1-17.

540 Schluter D (2001) Ecology and the origin of species. *Trends in Ecology and Evolution*, **16**, 372-380.

541 Schlüter PM, Xu S, Gagliardini V *et al.* (2011) Stearoyl-acyl carrier protein desaturases are associated  
542 with floral isolation in sexually deceptive orchids. *Proceedings of the National Academy of Sciences of*  
543 *the United States of America*, **108**, 5696-5701.

544 Sedeek KEM, Qi W, Schauer MA *et al.* (2013) Transcriptome and proteome data reveal candidate  
545 genes for pollinator attraction in sexually deceptive orchids. *PLoS One*, **8**, e64621.

546 Sedeek KEM, Scopece G, Staedler YM *et al.* (2014) Genic rather than genome-wide differences  
547 between sexually deceptive *Ophrys* orchids with different pollinators. *Molecular Ecology*, **23**, 6192-  
548 6205.

549 Sedeek KEM, Whittle E, Guthörl D *et al.* (2016) Amino acid change in an orchid desaturase enables  
550 mimicry of the pollinator's sex pheromone. *Current Biology*, doi: 10.1016/j.cub.2016.04.018.

551 Sheehan H, Moser M, Klahre U *et al.* (2016) *MYB-FL* controls gain and loss of floral UV absorbance, a  
552 key trait affecting pollinator preference and reproductive isolation. *Nature Genetics*, **48**, 159-166.

553 Shimizu KK, Kudoh H, Kobayashi MJ (2011) Plant sexual reproduction during climate change: gene  
554 function *in natura* studied by ecological and evolutionary systems biology. *Annals of Botany*, **108**,  
555 777-787.

556 Siepielski AM, DiBattista JD, Carlson SM (2009) It's about time: the temporal dynamics of phenotypic  
557 selection in the wild. *Ecology Letters*, **12**, 1261-1276.

558 Smadja CM, Canbäck B, Vitalis R *et al.* (2012) Large-scale candidate gene scan reveals the role of  
559 chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, **66**,  
560 2723-2738.

561 Sobel JM, Streisfeld MA (2015) Strong premating reproductive isolation drives incipient speciation in  
562 *Mimulus aurantiacus*. *Evolution*, **69**, 447-461.

563 Städler T, Roselius K, Stephan W (2005) Genealogical footprints of speciation processes in wild  
564 tomatoes: demography and evidence for historical gene flow. *Evolution*, **59**, 1268-1279.

565 Stankowski S, Sobel JM, Streisfeld MA (2015) The geography of divergence with gene flow facilitates  
566 multitrait adaptation and the evolution of pollinator isolation in *Mimulus aurantiacus*. *Evolution*, **69**,  
567 3054-3068.

568 Stankowski S, Sobel JM, Streisfeld MA (2017) Geographic cline analysis as a tool for studying genome-  
569 wide variation: a case study of pollinator-mediated divergence in a monkeyflower. *Molecular*  
570 *Ecology*, doi: 10.1111/mec.13645.

571 Stankowski S, Streisfeld MA (2015) Introgressive hybridization facilitates adaptive divergence in a  
572 recent radiation of monkeyflowers. *Proceedings of the Royal Society of London B: Biological Sciences*,  
573 **282**, 20151666.

574 Stinchcombe JR & Hoekstra HE (2008) Combining population genomics and quantitative genetics:  
575 finding the genes underlying ecologically important traits. *Heredity*, **100**, 158-170.

576 Streisfeld MA, Kohn JR (2005) Contrasting patterns of floral and molecular variation across a cline in  
577 *Mimulus aurantiacus*. *Evolution*, **59**, 2548-2559.

578 Streisfeld MA, Kohn JR (2007) Environment and pollinator-mediated selection on parapatric races of  
579 *Mimulus aurantiacus*. *Journal of Evolutionary Biology*, **20**, 122-132.

580 Streisfeld MA, Rausher MD (2009) Altered *trans*-regulatory control of gene expression in multiple  
581 anthocyanin genes contributes to adaptive flower color evolution in *Mimulus aurantiacus*. *Molecular*  
582 *Biology and Evolution*, **26**, 433-444.

583 Streisfeld MA, Rausher MD (2011) Population genetics, pleiotropy, and the preferential fixation of  
584 mutations during adaptive evolution. *Evolution*, **65**, 629-642.

585 Streisfeld MA, Young WN, Sobel JM (2013) Divergent selection drives genetic differentiation in an  
586 R2R3-MYB transcription factor that contributes to incipient speciation in *Mimulus aurantiacus*. *PLoS*  
587 *Genetics*, **9**, e1003385.

588 Tagu D, Colbourne JK, Nègre N (2014) Genomic data integration for ecological and evolutionary traits  
589 in non-model organisms. *BMC Genomics*, **15**, 490.

590 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.  
591 *Genetics*, **123**, 585-595.

592 Torrents D, Suyama M, Zdobnov E *et al.* (2003) A genome-wide survey of human pseudogenes.  
593 *Genome Research*, **13**, 2559-2567.

594 Turner TL, Bourne EC, Von Wettberg EJ *et al.* (2010) Population resequencing reveals local adaptation  
595 of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics*, **42**, 260-263.

596 van den Berg BHJ, Konieczka JH, McCarthy FM *et al.* (2009) ArrayIDer: automated structural re-  
597 annotation pipeline for DNA microarrays. *BMC Bioinformatics*, **10**, 30.

598 Venail J, Dell’Olivo A, Kuhlemeier C (2010) Speciation genes in the genus *Petunia*. *Philosophical*  
599 *Transactions of the Royal Society of London B: Biological Sciences*, **365**, 461-468.

600 Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of*  
601 *Sciences of the United States of America*, **106**, 9939-9946.

602 Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-  
603 with-gene-flow. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **367**,  
604 451-460.

605 Via S, Conte G, Mason-Foley C *et al.* (2012) Localizing  $F_{ST}$  outliers on a QTL map reveals evidence for  
606 large genomic regions of reduced gene exchange during speciation-with-gene-flow. *Molecular*  
607 *Ecology*, **21**, 5546-5560.

608 Visscher PM, Brown MA, McCarthy MI *et al.* (2012) Five years of GWAS discovery. *The American*  
609 *Journal of Human Genetics*, **90**, 7-24.

610 Volff JN (2005) Genome evolution and biodiversity in teleost fish. *Heredity*, **94**, 280-294.

611 Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype-phenotype map: the  
612 evolvability of complex organisms. *Nature Reviews Genetics*, **12**, 204-213.

613 Watt WB (2013) Specific-gene studies of evolutionary mechanisms in an age of genome-wide  
614 surveying. *Annals of the New York Academy of Sciences*, **1289**, 1-17.

615 Wayne ML, McIntyre LM (2002) Combining mapping and arraying: an approach to candidate gene  
616 identification. *Proceedings of the National Academy of Sciences of the United States of America*, **99**,  
617 14903-14906.

618 Weinreich DM, Delaney NF, DePristo MA *et al.* (2006) Darwinian evolution can follow only very few  
619 mutational paths to fitter proteins. *Science*, **312**, 111-114.

620 Widmer A, Lexer C, Cozzolino S (2009) Evolution of reproductive isolation in plants. *Heredity*, **102**, 31-  
621 38.



622 Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323-354.

623 Xu S, Schlüter PM (2015) Modeling the two-locus architecture of divergent pollinator adaptation:  
624 how variation in *SAD* paralogs affects fitness and evolutionary divergence in sexually deceptive  
625 orchids. *Ecology and Evolution*, **5**, 493-502.

626 Xu S, Schlüter PM, Grossniklaus U *et al.* (2012a) The genetic basis of pollinator adaptation in a  
627 sexually deceptive orchid. *PLoS Genetics*, **8**, e1002889.

628 Xu S, Schlüter PM, Schiestl FP (2012b) Pollinator-driven speciation in sexually deceptive orchids.  
629 *International Journal of Ecology*, **2012**, 285081.

630 Xu S, Schlüter PM, Scopece G *et al.* (2011) Floral isolation is the main reproductive barrier among  
631 closely related sexually deceptive orchids. *Evolution*, **65**, 2606-2620.

632

633

## **BOX 1**

### **Box 1A: Applying a genomic view to existing molecular studies of adaptation**

**Wing colour patterning in *Heliconius* butterflies:** Butterflies in the genus *Heliconius* show extensive Müllerian mimicry in wing patterning, particularly between the distantly related *H. erato* and *H. melpomene*, which have multiple shared wing pattern variants (Kronforst & Papa 2015). Purifying selection within species for the presence of red on the forewing has been demonstrated, as has selection against foreign morphs, demonstrating the importance of wing patterning (Kronforst & Papa 2015). Starting from a genomic region identified in previous mapping experiments, Reed *et al.* (2011) used comparative transcriptomics to identify *optix*, a homeobox transcription factor, as the candidate gene underlying the red colour patterns in multiple *Heliconius* species. Hybrid butterflies showed strong association between *optix* sequence variation and colour pattern (Reed *et al.* 2011); within *H. melpomene*, colour races demonstrated elevated  $F_{ST}$  in the genomic region containing *optix* (Nadeau *et al.* 2012). Comparative developmental evidence suggests that the co-option of *optix* as a driver of colour patterning is a recent novelty within the Heliconiini (Martin *et al.* 2014).

**Reproductive isolation between *Mimulus lewisii* and *M. cardinalis* (monkeyflowers, Box Figure I):** In sympatry, 97% of the reproductive isolation between these two species is mediated by their animal pollinators (Ramsey *et al.* 2003). Quantitative trait loci for traits influencing pollinator behaviour (*e.g.* anthocyanin and carotenoid pigmentation, Bradshaw *et al.* 1995; Bradshaw *et al.* 1998; Bradshaw & Schamske 2003; and floral scent, Byers *et al.* 2014) co-localize with those affecting pollinator efficiency (*e.g.* stamen and style length, Bradshaw *et al.* 1995; Bradshaw *et al.* 1998; Fishman *et al.* 2013; and corolla length, Fishman *et al.* 2013) and pollen viability (Fishman *et al.* 2013). These clusters of loci affecting reproductive isolation between the two species are located within putatively rearranged regions of the genome (Fishman *et al.* 2013). Although it is unclear whether genome

evolution processes initially drove reproductive isolation in this system, recombination suppression has served to maintain suites of traits maintaining isolation in sympatry.

**Reproductive isolation between *Petunia axillaris* and *P. exserta*:** The narrow endemic *Petunia exserta* hybridizes with the closely related *P. axillaris* due to pollen transfer by hummingbirds (Lorenz-Lemke *et al.* 2006). Several loci underlying traits responsible for differences in pollinator behaviour and efficiency have been identified in multiple *Petunia* species, *e.g.* scent (Klahre *et al.* 2011), anthocyanins (Hoballah *et al.* 2007) and UV absorbance (Sheehan *et al.* 2016), corolla area and tube length (Venail *et al.* 2010), and pistil and stamen length (Hermann *et al.* 2013). At least three loci representing five of these traits have been found to be in tight linkage in a region with potentially suppressed recombination between these two species (Hermann *et al.* 2013); as with *Mimulus*, this suggests a role for genome evolution underlying reproductive isolation between these two species, although the processes driving genome evolution are as yet unclear in this system. Further characterization of these loci and any role of genomic evolution is needed; in particular, as with *Mimulus*, a genome-wide scan for signatures of selection by pollinators in the field would be a valuable addition to this system.

**Box 1B: Insights into genome-wide approaches from existing orthologous gene annotations**

**Cryptic coloration in *Timema cristinae* stick insects (Box Figure II):** Two host plant races of *T. cristinae* differ in their pigmentation patterning, which is under divergent selection, yet nonetheless maintain gene flow (Nosil *et al.* 2002; Nosil 2007). A third, melanistic morph shows mating advantages, increased pathogen resistance, and can maintain crypsis in stem microhabitats (Box Figure II, central panel); modelling shows these melanistic individuals may be maintaining gene flow between the two races (Comeault *et al.* 2015a). Genome-wide association mapping found a single locus responsible for 95% of the phenotypic variation in melanism; the most significant SNPs mapped

681 to a putative cysteinyl-tRNA synthetase, an enzyme known to be required for melanin synthesis  
682 (Comeault *et al.* 2015a). Further functional work will be required to verify the function of this gene in  
683 *Timema* but will allow more in-depth work on melanism in this and other *Timema* species (*e.g.*  
684 Comeault *et al.* 2015b).

685 **Serpentine soil adaptation in *Arabidopsis lyrata*:** Plants locally adapted to serpentine soils must  
686 cope with increased heavy metal concentrations and poor nutrient availability. Turner *et al.* (2010)  
687 sequenced whole genomes from *A. lyrata* plants from serpentine and non-serpentine soils to locate  
688 SNPs segregating with soil type, performed a sliding window  $F_{ST}$  scan to search for signatures of  
689 selection on serpentine soils, and scanned the genome for copy number variants. Many of the  
690 segregating SNPs and outliers corresponded to annotated metal ion transporters from the closely  
691 related *A. thaliana* genome, and serpentine plants appeared to have multiple duplications of regions  
692 containing toxin-processing loci. The authors argue for the role of many distinct loci in serpentine soil  
693 adaptation in this system.

694 **Adaptation to anthropogenic metal pollution in *Perca flavescens* (yellow perch):** Yellow perch  
695 subject to anthropogenic metal pollution have evolved a rapid life history strategy to compensate for  
696 metal toxicity (Couture & Pyle 2008). Using markers derived from existing whole transcriptomic data,  
697 the genome was scanned for outliers in fish from contaminated and clean lakes, resulting in three  
698 markers under directional selection. The single nuclear genome marker corresponds to the cyclin G1  
699 gene (tentatively identified via orthologues in four separate fish species), which is strongly associated  
700 with metal levels in the fish, contains a SNP at a position highly conserved across vertebrates, and  
701 may be associated with increased growth rates, thus arguing for it as a target of selection (Bélanger-  
702 Deschênes *et al.* 2013). Future work characterizing cyclin G1 in yellow perch will help shed light on its  
703 role in adaptation to metal pollution.

In all of these examples, functional characterization and validation of candidate genes is yet to be done.

**Box 1C: Integrating genetic and genomic adaptation and ecological speciation: case studies**

**Cryptic coloration in *Peromyscus* mice:** Wild populations of two species of *Peromyscus* mice display recently-evolved variation in pigmentation that correlates with habitat (substrate), with pale mice found on pale substrate and dark mice found on dark substrate; the matching of mouse and substrate colour is under selection by predators (Hoekstra 2006). In beach mice (*P. polionotus*), *Mc1r*, the melanocortin-1 receptor gene, is responsible for this difference, with a single SNP displaying a strong genotype-phenotype association in multiple body regions; extensive functional study supports this link, and allele frequencies segregate with the dominant colour phenotype across a large population range (Hoekstra *et al.* 2006). In the deer mouse (*P. maniculatus*), a similar environment-phenotype match occurs, but the responsible gene is instead *Agouti*, the product of which is an inverse agonist of *Mc1r*. *Agouti* has multiple SNPs between light and dark mice, and displays a loss of nucleotide variation characteristic of a selective sweep (Linnen *et al.* 2009). Further work in this species using genomic techniques – albeit applied only to SNPs within this locus rather than genome-wide – verified the role of *Agouti* in melanism, and has shown that individual SNPs within *Agouti* are responsible for pigmentation differences in multiple body regions, that these SNPs are under selection, and that the “dividing up” of body regions by SNP allows for minimal pleiotropic effects (Linnen *et al.* 2013). Alternative transcript processing of *Agouti* appears to underlie the light-dark transition in these mice, with one untranslated exon specific to the hair cycle showing signatures of positive selection (Mallarino *et al.* 2017, this issue).

**Reproductive isolation in sexually deceptive *Ophrys* orchids (Box Figure III):** *Ophrys* flowers resemble female insects and are pollinated by males searching for mates; pollinator identity is typically species-specific and is the main driver of reproductive isolation between *Ophrys* species (Xu

*et al.* 2011; Sedeek *et al.* 2014). A variety of floral phenotypes, particularly hydrocarbons on the floral surface that mimic female insect pheromones, are important for pollinator specificity. Pollinator-relevant species-specific differences in alkene hydrocarbon double bond position are due to differences in several stearyl-acyl carrier protein desaturases (SADs), which have been functionally characterized in this system (Schlüter *et al.* 2011; Sedeek *et al.* 2016). This complete functional characterization, together with the estimation of fitness effects of different hydrocarbon phenotypes (Xu *et al.* 2012a), has allowed the modelling of various evolutionary trajectories *via* changes in SAD function (Xu & Schlüter 2015). Additional candidate genes for other floral phenotypes were also identified via transcriptome and proteome analyses (Sedeek *et al.* 2013). Genome-wide work found only weak species differentiation, and outlier scans identified only a very small amount of the genome (including candidate hydrocarbon biosynthetic genes) as associated with this differentiation. Together, this suggests that genic, rather than genomic, speciation is at work in this system (Sedeek *et al.* 2014).

**Reproductive isolation in the *Mimulus aurantiacus* species complex:** Red-flowered *Mimulus aurantiacus* ssp. *puniceus* and yellow-flowered *M. aurantiacus* ssp. *australis* are parapatric and are undergoing incipient ecological speciation associated with pollinator-mediated reproductive isolation in the contact zone (Streisfeld & Kohn 2007; Sobel & Streisfeld 2015). Detailed and thorough molecular work identified a single R2R3 MYB transcription factor, MaMyb2, which drives this divergence (Streisfeld & Rausher 2009; Streisfeld *et al.* 2013). Analysing allelic differentiation and clinal shape at this locus as well as across the genome (Streisfeld & Kohn 2005) demonstrated that MaMyb2 is under selection (Streisfeld *et al.* 2013). Genome-wide analysis of differentiation argued against allopatric speciation with secondary contact, supporting the parapatric speciation hypothesis and suggesting differentiation only at locally adaptive genomic regions (Stankowski *et al.* 2015). In this issue, Stankowski *et al.* (2017) combine clinal analysis and a genome-wide scan for differentiation between the two subspecies, concluding that a set of 130 loci have clines with a similar shape to

753 clines in floral traits, indicating that they – or loci they are linked to – are involved in pollinator  
754 isolation. In the hybrid zone between the two subspecies, outlier loci are distributed across the  
755 genome, in contrast to the concept of genomic islands of speciation (Stankowski *et al.* 2017).

756

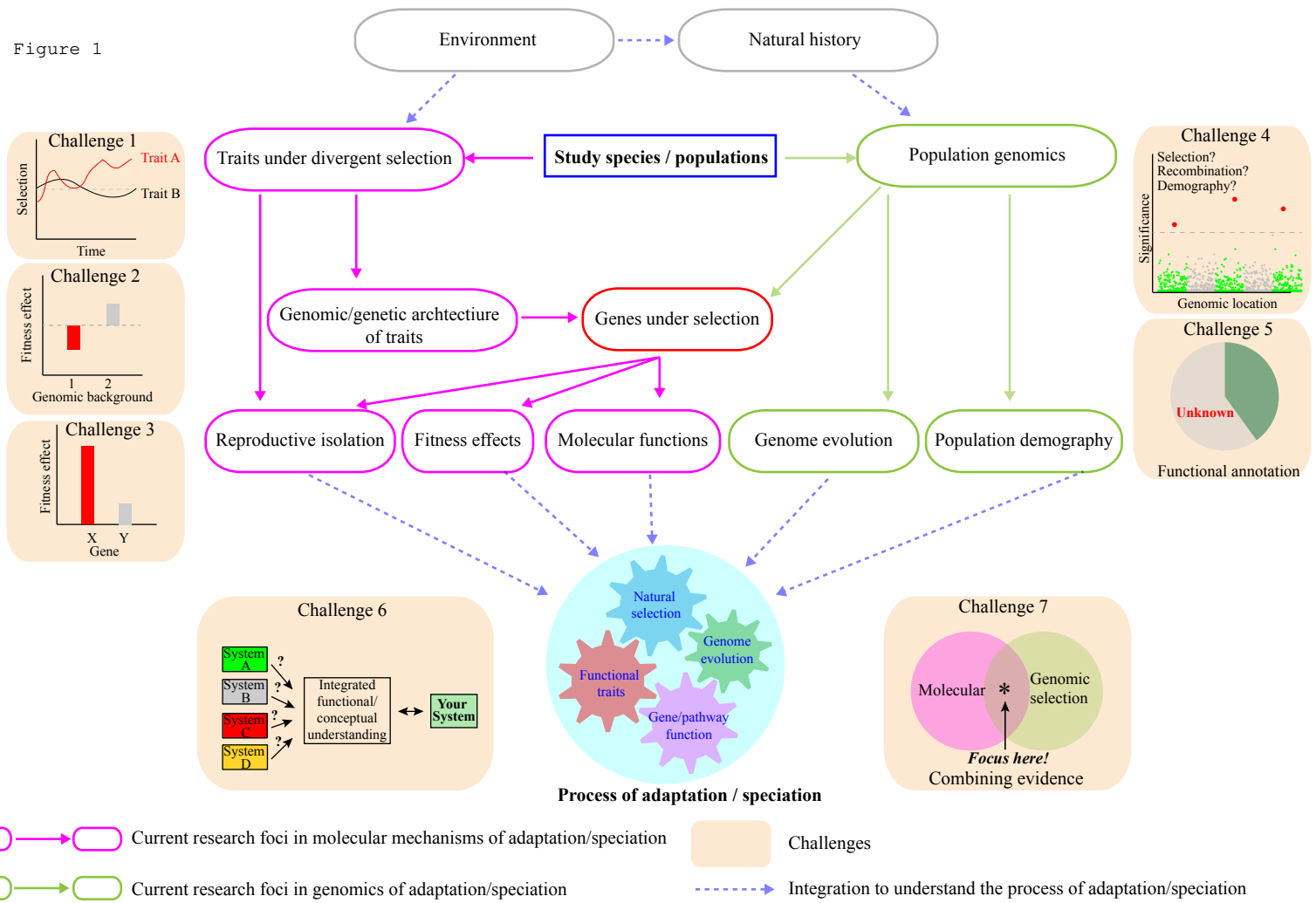
757

## FIGURE 1

Current research foci and challenges in studying adaptation and speciation using molecular (magenta arrows) and genomic (green arrows) approaches. Light red boxes highlight the challenges in these research areas, as follows: Challenge 1, selection on traits might be dynamic and dependent on time; Challenge 2, the fitness effects of individual genes might be dependent on genomic background, such as epistatic interactions among genes; Challenge 3, different genes associated with traits that affect fitness might have different fitness effects, and these effects might additionally vary with environmental conditions; Challenge 4, candidate loci identified using adaptation/speciation genomics might be spurious, with significant outliers resulting from demographic changes, false positives, or selective sweeps, rather than selection on the locus itself; and Challenge 5, even in model organisms (*e.g. Arabidopsis thaliana*, *Drosophila melanogaster*, or *Mus musculus*), the function of a large proportion of genes remains unknown. Two remaining challenges of an integrative approach – Challenge 6, synthesizing results from multiple systems into a conceptual whole; and Challenge 7, combining evidence from molecular and genomic approaches – must also be addressed. When addressing these challenges, it will be important to keep several points in mind: first, the importance of the environment is paramount, and *in natura* approaches should be used where possible; second, in some cases it may be possible to harness natural variation, particularly in less tractable systems; and third, it will remain critically important to functionally validate any candidate genes resulting from these approaches. The bottom circle depicts the integration of molecular and genomic approaches to studying adaptation and speciation, with some suggested future integrative approaches (blue dashed arrows). By linking natural selection, genome evolution, functional traits, and gene and pathway function, we can achieve an integrated research programme.



Figure 1



782    **BOX FIGURES**

783    **Box Figure I:** *Mimulus lewisii* (top) and *M. cardinalis* (bottom), sister species reproductively isolated  
784    largely by pollinator preference for divergent floral traits. Genes and QTL underlying these floral traits  
785    are clustered in areas of recombination suppression, maintaining distinct floral phenotypes. All  
786    images by Kelsey Byers.

787    **Box Figure II:** *Timema cristinae* stick insects. The *Ceanothus* (left) and *Adenostoma* (right) ecotypes  
788    differ in the presence of a white dorsal stripe, aiding camouflage on their different host plants.  
789    Melanistic *T. cristinae* (centre) are able to hide on the stems of both host plants, serving as a bridge  
790    for gene flow between ecotypes. Images courtesy Aaron Comeault (left, right) and Moritz Muschick  
791    (centre), reproduced with permission.

792    **Box Figure III:** Three species of sexually deceptive orchids in the genus *Ophrys* (left to right, *O.*  
793    *garganica*, *O. exaltata* subsp. *archipelagi*, and *O. sphegodes*). *Ophrys* orchids mimic the sex  
794    pheromones of their bee pollinators, which ensures specificity in pollen transfer and provides an  
795    early-acting reproductive barrier. Species divergence appears to be largely genic rather than  
796    genomic, in agreement with their position in the early stages of the speciation process. All images by  
797    Philipp Schlüter.

798











